DOCUMENT RESUME

ED 412 825 HE 030 616

AUTHOR	Harris, Jimmy Carl; Clemmons, Sandra
TITLE	Utilization of Standardized Critical Thinking Tests with Developmental Freshmen.
PUB DATE	1996-10-00
NOTE	11p.; Paper presented at the National Conference on Research in Developmental Education (2nd, Charlotte, NC, October 23-26, 1996).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	College Entrance Examinations; College Freshmen; Compensatory Education; *Critical Thinking; *Developmental Studies Programs; Factor Analysis; Higher Education; Placement; Predictive Measurement; *Remedial Programs; *Screening Tests; Test Construction; Test Reliability; *Test Theory; Test Validity; Transitional Programs
IDENTIFIERS	*California Critical Thinking Skills Test (College); *Watson Glaser Critical Thinking Appraisal

ABSTRACT

This paper presents the results of a search for an appropriate test of critical thinking to screen college freshmen. The search for an appropriate test of critical thinking was initiated in the Fall 1995 semester at an open-admissions comprehensive university, which normally assigns entering freshmen with ACT composite scores of 17 or less to remediation programs for English, mathematics, and reading, in addition to a 3-semester hour critical thinking course. Two tests, the Watson-Glaser Critical Thinking Appraisal and the California Critical Thinking Skills Test were administered to 27 and 32 students, respectively, enrolled in developmental education courses. The study sought to determine: (1) if either test could serve as a predictor or course performance; (2) what relationship existed between test content and course content; (3) how local scores on the tests compared to national norms; and (4) were the test versions equivalent. Results showed that differences in score means for the pre- and post-tests were statistically insignificant for both tests, suggesting that the remediation course had little impact on student performance, concluding that the tests, at best, have problematical usefulness as a predictor for placement purposes. Appended are four tables that define terms and summarize data. (BF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 412 825

UTILIZATION OF STANDARDIZED CRITICAL THINKING TESTS WITH DEVELOPMENTAL FRESHMEN

Dr. Jimmy Carl Harris
Assistant Professor of Developmental Critical Thinking
Southeastern Louisiana University

Mrs. Sandra Clemmons
Graduate Student
Southeastern Louisiana University

A paper presented at the Second National Conference
on Research in Developmental Education

Charlotte, North Carolina

23-26 October, 1996

BEST COPY AVAILABLE

IAE 030 616

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY

Jimmy Carl Harris

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

UTILIZATION OF STANDARDIZED CRITICAL THINKING TESTS WITH DEVELOPMENTAL FRESHMEN

Introduction

This study took place in an open-admissions, comprehensive university with an enrollment of approximately 14,000. The university assigns freshmen with ACT composite scores of 17 or less to the Department of Developmental Education; in recent years, this has amounted to about one-third of the entering freshmen. The department offers remediation in English, mathematics, and reading, as well as a three semester-hour critical thinking course that is required of all developmental students (Table 1). The department has placement tests, used in conjunction with ACT subscores, for English, mathematics, and reading. However, there is no placement test for critical thinking.

To aid in the search for an appropriate placement test, a study was initiated during the Fall, 1995 semester, when a number of standardized critical thinking tests were considered for study. The final choices were limited to tests that had two versions, necessary for the pre-test/ post-test design of the study, and were commercially available. Also, because of the anticipated use of one of the tests for placement purposes, meaning that large numbers of tests would be administered and scored in a relatively short period, tests were sought that could be administered in no more than one hour and could be easily scored. With these criteria in mind, the Watson-Glaser Critical Thinking Appraisal (WGCTA) and the California Critical Thinking Skills Test (CCTST) were chosen for study.

The Tests

Watson-Glaser Critical Thinking Appraisal. Forms A and B of the WGCTA were first copywrited in 1951 (Gibbs, 1985); the versions utilized for this study were copywrited in 1980 (Watson & Glaser, 1980). Norris & Ennis (1989) observed that it is probably the most widely used critical thinking test and often serves as the standard for comparison when studies of such tests are conducted. They also stated that "information on the test's validity includes studies which show increases in test performance following instruction in critical thinking, and correlations of the test with measures of general intelligence, aptitude, and achievement" (p. 61). The WGCTA has five sections of sixteen items each, as described in Table 2. In the manual accompanying the tests, Watson and Glaser (1980) state that "although the *Critical Thinking Appraisal* is intended as a test of power and speed, a 40 minute period of working time can be imposed for the sake of convenience in administration" (p. 2).

California Critical Thinking Skills Test. Copywrited in 1992 and updated in 1994 (Facione & Facione, 1994), the CCTST is the much more recent of the tests studied. For this reason, the literature is largely limited to reports from the publishers. The CCTST contains 34 items, each of which apply to more than one of the subscales listed in Table 3. In the instructions manual, the publishers recommend that the test-takers be given "45 minutes (unless

you have decided to extent the time for some reason and are planning to develop local norms)" (p. 6-7).

Design of the Study

For this study, answers to four questions, relating to both tests, were sought:

> **Would either test serve as a predictor of course performance?** If either test proved to be a good predictor, then perhaps it could be used as a placement test, making it possible for incoming developmental students to "test out" of critical thinking. Prediction was based on the extent to which the students above the median on the pre-test of each test were also above the median on their overall course grade.

> **What was the relationship between test content and course content?** By considering the course to be the treatment in a pre-test (version A)/post-test (version B) design, more could be known about the extent to which the tests measured the course content.

> **How would the local scores on the tests compare to national norms provided by the test publishers?** A major goal of the researchers was to find a test appropriate for *developmental* students, but the tests appeared to be normed for non-developmental students.

> **Were the test versions equivalent?** The literature contained concerns about test version equivalency, especially in the case of the CCTST.

Four Developmental Critical Thinking sections provided the subjects. The race, gender, and academic level of the subjects was representative of the university's developmental students as a whole, including an ACT range of 13 - 17. In order to minimize the number of variables, all of the sections were taught by the same professor, using the same lesson plans, course materials, and course evaluations. To ensure that all of the subjects had ample time to respond to all test items, 60 minutes were allowed for taking the tests.

Results

During the Spring, 1996 semester, data were collected (Table 4). A comparison of pre-test scores to final course grades (for those who both took the pre-test and finished the course) was made. 50% of those who placed above the median on the CCTST pre-test, and 61.5% of those who placed above the median on the WGCTA pre-test, also finished the course above the median for their section.

The difference of score means (CCTST N = 32, WGCTA N = 27) for the pre- and post-tests were, for both the CCTST and the WGCTA, statistically insignificant, suggesting that the course had little impact on student performance on these tests. For each test, N represents two combined sections, with only those subjects that took the pre-test and the post-test and

finished the course.

A comparison with the publishers' test norms was also made. The CCTST subjects' pre-test mean score (10.9, N = 43) was significantly below the mean score norm (15.89, N = 781) reported by the CCTST publisher, but this norm, while it does represent students who had not taken a critical thinking test, appears to be based on testing in but one university, and does not appear to include developmental students. On the WGCTA pre-test, the subjects' mean score (44.5, N = 37) was significantly less than the norm means for community college students (51.9, N = 388) and four-year college freshmen (53.8, N = 824), as reported by the publishers. The WGCTA publishers do not reveal the number of norm subjects who may have taken a critical thinking course, or how many may have been developmental.

In addition, doubts about test version equivalency were raised, especially with the CCTST. Of the CCTST subjects who finished above the median, in terms of final course grade, 57.14% had post-test scores that were lower than, or equal to, their pre-test scores. These results reflect a concern raised in the literature (Jacobs, 1994). The WGCTA subjects fared somewhat better; 30.77% of those above the median on course grades had lower post-test scores. It should be noted that both publishers go to some length to present their respective "A" and "B" versions as equivalent.

Discussion

With students and a course such as those included in this study, using either of these tests as a predictor suitable for placement purposes would be problematical, at best. With only half (CCTST) or slightly more than half (WGCTA) of those above the median on the pre-test also in the top half in terms of course grades, the notion that doing well on the pre-test indicates a likelihood of doing well in the course is reduced to about the same odds as a coin-toss.

In his review of 27 studies of critical thinking instruction in higher education, McMillian (1987) concluded that "what is lacking in the research is a common definition of critical thinking, good instrumentation to provide specific measurement, and a clear theoretical description of the nature of an experience that should enhance critical thinking" (p. 3). While the studies reviewed by McMillan covered a much broader range of issues than this study, his conclusions seem appropriate both there and here. The critical thinking course studied here was, in genesis, heavily informed by the widely read and authoritative work of John Chaffee (1991) and, especially, Richard Paul (1993). However, a look at Tables 1, 2, and 3 establishes that, while there are content similarities between this course and each of the two critical thinking tests, neither test appears to be a good fit for the course. These differences may not indicate any particular strength or weakness in the course or in either test. But, when these content differences are viewed along with the closeness of the pre-test/post-test scores, there emerges the rather clear impression that, in this instance, neither one of these two standardized critical thinking tests is appropriate for a developmental course based on the work of leading authorities in the field. A commonly-accepted definition of critical thinking and, therefore, a widely accepted means of measuring it are not yet in hand.

The significantly sub-norm scores of these subjects on both tests gave rise to another concern about the appropriateness of either test for these developmental students. With scores below 70% (WGCTA) and 60% (CCTST), it is difficult to claim that any of the subjects in this study displayed mastery on either pre-test. When added to the problems with content matching and the lack of test version equivalency, this lack of mastery becomes an even greater concern. Without mastery, there is little justification for claiming that any test could be used to predict success; without this prediction, there is scant rationale for using a test for placement purposes.

A further concern, encountered in the literature and in this study, has to do with the publishers' claims of test version equivalency. Jacobs (1994) found that the two versions of the CCTST did not seem to be equivalent, and this study seems to support that conclusion. While Jacobs did a much more comprehensive job of research than was attempted in this study, the experiences were about the same: students performed less well on version B than on version A, and a lack of equivalency seemed to be indicated. A particularly troublesome statistic produced by this study was the 57.14% of the top-half students who actually did less well on the post-test than on the pre-test. This study produced less evidence of equivalency problems with the WGCTA than with the CCTST, but 30.77% of the WGCTA cohort also scored lower after the course than before.

Conclusions and Implications

Because of the results of this study, the department has placed in temporary abeyance any decisions about using these tests with developmental students. Clearly, the study, as originally conceived, did not produce results that would engender confidence in using either test for placement purposes. However, the issue remains important to the university; for that reason, more data, based on sections taught by several instructors, are being collected. This will provide a broader base for studying the correlation between course and test content. Other matters to be considered include appropriate local norms for both tests, possible contribution of these norms to national norms, and the utility of these or other instruments for placement purposes. These expanded studies will serve as an important contribution to the developmental education literature.

Table 1

Elements and Standards of Critical Thinking

These elements and standards are drawn from the work of Richard Paul (1993). They form the conceptual foundation for the critical thinking course in this study.

Elements of Critical Thinking
Purpose
Questions
Concept
Information
Assumption
Point of View
Interpretation
Inference
Conclusion
Implication
Consequence

Standards of Critical Thinking
Clarity
Accuracy, Precision, and Specificity
Relevance and Significance
Breadth, Depth, and Completeness
Fairness and Consistency
Logic and Justifiability

Table 2

Watson-Glaser Critical Thinking Analysis (WGCTA) Sub-tests

Responses to the 80 items on this test may be evaluated in terms of the five sub-tests, each of which contains 16 items.

Sub-test	WGCTA Meaning
Inference	Discriminating among degrees of truth or falsity of inferences drawn from given data.
Recognition of Assumptions	Recognizing unstated assumptions or presuppositions in given statements or assertions.
Deduction	Determining whether certain conclusions necessarily follow from information in given statements or premises.
Interpretation	Weighing evidence and deciding if generalizations or conclusions based on the given data are warranted.
Evaluation of Arguments	Distinguishing between arguments that are strong and relevant and those that are weak or irrelevant to a particular question at issue.

Table 3

California Critical Thinking Skills Test (CCTST) Sub-scales

Responses to the 34 items on this test may be evaluated in terms of three sub-scales based on *The Delphi Report*, or in terms of two sub-scales based on traditional categories.

Delphi Sub-scale	CCTST Meaning
Analysis	Examining ideas identifying arguments Analyzing arguments
Evaluation	Assessing claims Assessing arguments
Inference	Querying evidence Conjecturing alternatives Drawing conclusions

Traditional Sub-scale	CCTST Meaning
Deductive Reasoning	The assumed truth of the premises purportedly necessitates the truth of the conclusion.
Inductive Reasoning	An argument's conclusion is purportedly warranted, but not necessitated, by the assumed truth of its premises.

Table 4

Data

Watson-Glaser Critical Thinking Appraisal

N = 27

Course Grades: mean = 81.98%, median = 83.5%, range = 62% - 93%.

Pre-test: mean = 55.67%, median = 54%, range = 41% - 69%.

Post-test: mean = 58.96%, range = 44% - 74%.

A *t*-test of pre-test and post-test means found that their differences were not statistically significant at the .05 level.

California Critical Thinking Skills Test

N = 32

Course Grades: mean = 81.68%, median = 81%, range = 66.5% - 95%.

Pre-test: mean = 32.19%, median = 29%, range = 12% - 56%.

Post-test: mean = 33.22%, range = 18% - 47%.

A *t*-test of pre-test and post-test means found that their differences were not statistically significant at the .05 level.

Sources Cited

Chaffee, J. (1991). Thinking critically (3rd ed.). Boston: Houghton Mifflin.

Facione, P. & Facione, N. (1994). Test manual: The california critical thinking skills test. Millbrae, CA: The California Academic Press.

Gibbs, L. (1985). Teaching critical thinking at the university level: A review of some empirical evidence. Informal Logic, 7(2), 137-149.

Jacobs, S. (1994, May). Technical characteristics and some correlates of the california critical thinking skills test. Paper presented at the Annual Forum of the Association for Institutional Research, New Orleans, LA.

McMillan, J. H. (1987). Enhancing college students' critical thinking: A review of studies. Research in Higher Education, 26 (1).

Norris, S. & Ennis, R. (1989). Evaluating critical thinking. Pacific Grove, CA: Midwest Publications.

Paul, R. (1993). Critical thinking: How to prepare students for a rapidly changing world. Santa Rosa, CA: Foundation for Critical Thinking.

Watson, G. & Glaser, E. (1980). Manual: The watson-glaser critical thinking appraisal. New York: Harcourt Brace Jovanovich.

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor: NOT APPLICABLE
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name: NOT APPLICABLE
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

Higher Education
TheGeorge Washington University
One Dupont Circle, NW, Suite 630
Washington, DC 20036-1183

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2d Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

(Rev. 6/96)